Analyzing Science Operations for the Search for Life as part of a Multi-Year Robotic Campaign to Explore the Atacama Desert

Second Year Report

Sponsor: Applied Information Systems Research Program,
National Aeronautics and Space Administration
Project Number NNG05GA51G

Geb Thomas, PI
Nathalie Cabrol, NASA Ames
David Wettergreen, CMU
Robert Anderson, JPL
Peter Coppin, CMU
Edmond Grin, Ames

Contact Information
Geb Thomas
Department of Mechanical and Industrial Engineering
The University of Iowa
Iowa City, IA 52246
Phone: (319) 335-5936
Fax: (319) 335-5669
URL: http://grok.ecn.uiowa.edu

August 17, 2006

Summary

This second annual report updates the first annual report submitted in February 2006. Since the last report we largely concluded the analysis from the October 2005 LITA experiment, prepared our findings for publication, and are developing new models to frame our next year's effort. The principle result from our efforts so far has been the development of a new "Directed Autonomy" model of scientists' interaction with exploration robot. Pending peer review, the new model is intended to supplant the Supervisory Control model, which our evidence suggests fails to account for the behavior patterns observed in the LITA experiment. The principle difference between the models is that Supervisory Control suggests that the scientists' tools should be focused towards understanding and monitoring the state of the robot whereas the Directed Autonomy model suggests that the tools should emphasize the interpretation of and learning from the data provided by the robot. We have also developed a preliminary mathematical model describing how the design of a camera system may assist scientists in using a robot to identify features of interest in the environment. Finally, we are currently completing a paper to be prepared in a special issue of the Journal of Geophysical Research with the other science reports from the LITA field test. Our contribution will be to explain the challenges of evaluating the scientific success of a robotic field test and to describe three different techniques that we employed in this field test.

Although the project has enjoyed many successes, we also discovered that there are fundamental challenges to our paradigm. Our presumption that the science team would work by forming and testing hypotheses about the remote environment has not been well supported by our observations. We found that scientists often hesitated to state their observations as hypotheses and that in the field there was substantial disagreement and inconsistency about whether statements we interpreted as specific, individual hypotheses were correct or incorrect.

Rather than forming a linear, incrementally progressing structure, the scientists' thinking patterns seem to be more like an evolving network of ideas and beliefs. This is important to our research because we wanted to trace the effect of individual data sources on the formation of scientific hypotheses and observations. If we cannot isolate and determine the outcome of particular hypotheses, it is difficult to precisely map the flow of information that generated that hypothesis. In order to continue, we need to modify our paradigm to work with an evolving network of ideas. Consequently, our research is now faced with the challenge of understanding how to: capture this evolving network of ideas, evaluate the network, measure the effect the robot may have in shaping this network, and developing tools to assist the scientists in building and forming appropriate conclusions from the network.

This summer the interdisciplinary Obermann Center at the University of Iowa invited Thomas and two Iowa geologists, Ingrid Ukstins Peate and Mark Reagan, to spend a month in sequestered, collaborative research. Since Ukstins Peate had participated in

the ground truth evaluation of the LITA mission and Reagan had participated in a 1999 robotic expedition to Silver Lake, Arizona, all of the team members were familiar with the many challenges of evaluating rover field tests. This interdisciplinary group developed a formal model based on semiotic theory to describe a geologist's evolving thinking about an environment. Peter Coppin is developing an interface that will allow us to test the efficacy of the new approach. Rather than develop a monolithic field test for the third year, we have decided to develop a series of focused data sets and test them with students, classes of students, and professional geologists.

Introduction

In the austral spring (September-October) of both 2004 and 2005, a multi-institutional, multi-national effort led by Carnegie Mellon University conducted an extended duration robotic exploration of the Atacama Desert in Chile to develop technologies and methods for upcoming NASA missions that will search for evidence of past life on Mars. Our project leverages this large-scale effort to measure and improve the effectiveness of robotic science operations. This research will extend current analysis of rover-mediated geology to rover-mediated habitat characterization. The work emphasizes the effects of different data collection and display techniques on the science team's conclusions. The principal hypothesis of this research is that the quality and reliability of scientific conclusions regarding past or present life in arid environments is dependent on the type of evidence collected by the rover, the scientists' data analysis techniques, the processes used by the scientists' to form and share hypotheses and conclusions, and the science operations software. The principal objectives of this research, each specifically associated with AISR program objectives, are to: 1) reduce mission development time by analyzing how scientists characterize a habitat, 2) reduce mission development risk by identifying mission-critical and problematic analysis tasks, 3) increase science return from the data by analyzing long-traverse science collection strategies, and 4) increase data return by refining the science interface to improve analysis effectiveness.

This research will analyze the processes used by astrobiologists and geologists when searching for signs of life in a Mars-like environment. Our previous and ongoing work with robotic geology has successfully characterized limitations in scientific interpretations caused by rover sensors, differences in scientists' interpretations, and limitations of the science interface. These limitations were identified and studied using perceptual experiments in which scientists analyzed sample images and physical specimens. In addition, transcripts of scientists participating in a simulated rover field experiments have been examined to further understand these limitations. This project will quantify analyst and instrument limitations that could affect the success of future missions in the search for life on Mars and will develop mitigating strategies to avoid inappropriate conclusions regarding the presence of life on Mars.

Background

The science information interface that displays rover-collected data during a planetary exploration mission is the science team's principal window to another planet and is therefore critical to the mission's success. Previous rover missions and field tests have demonstrated that the design of both the rover's instruments and the science information interface affect the quality of the scientific interpretation of the remote environment. Our recent research has isolated and quantified specific differences in scientific interpretation between information presented in a picture and information directly perceived. What is not yet established, however, is the degree to which limitations in specific science analysis tasks can affect the scientific success of a rover mission and whether strategies to mitigate these limitations can significantly improve scientific success. This is a critical research issue because if problems associated with rover-mediated scientific interpretation are not identified and corrected, they could mislead time-stressed mission scientists toward overconfident, under confident, or erroneous scientific conclusions. These are serious problems that threaten to jeopardize NASA's success in robotic planetary exploration.

The long-term goal of this research is to improve the effectiveness of rover operations by identifying the operational limits of scientific interpretation and discover means to eliminate these limitations. The objective of this research project, which is the next step towards attaining our long-term goal, is the analysis of the tasks conducted by a science team during a rover mission and the refinement of their science information interface to enhance science operations. Specifically, we are analyzing the operations of a science team engaged in a field campaign in the Atacama Desert, characterizing which analysis tasks are the most limiting to the success of this campaign, designing and implementing solutions that eliminate these limitations, and confirming that the solution may be generalized to other remote exploration experiences. We need to carefully and exhaustively characterize the analysis tasks conducted by the science team in order to isolate the connection between any misleading or erroneous scientific conclusions and the details of the analysis of the original rover data that led to the misconception. Once the connection between the analysis and the misleading conclusion has been made, techniques may be developed that reduce the chance that similar mistakes will be repeated in the future. This research will lead to rover science operation software that mitigates common and egregious interpretation errors and improves the efficiency of time-consuming tasks. Our research team is particularly qualified to conduct this research because of our extensive experience building rover science interfaces, building exploration rovers, running rover field experiments, participating in planetary missions, and exploring remote and planetary terrains.

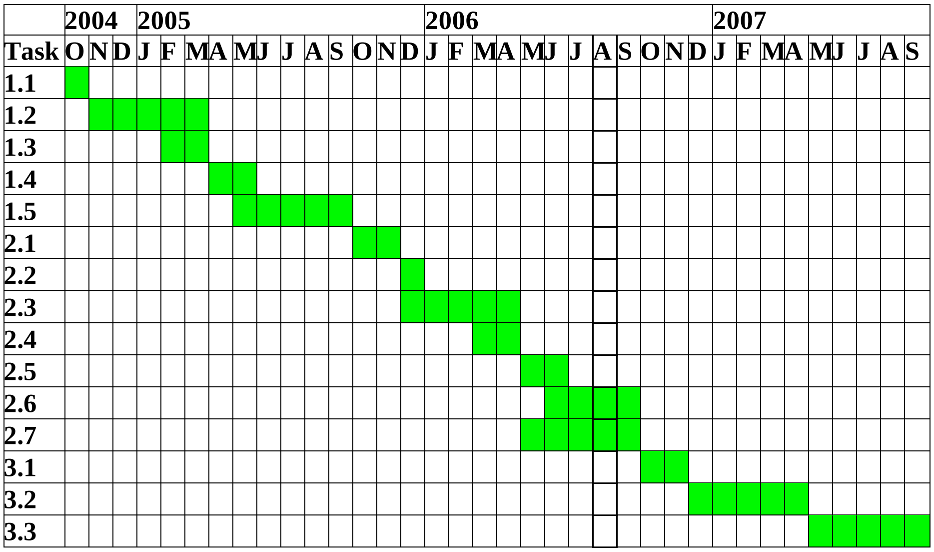
2.1 Objectives and Expected Significance

1. **Reduce mission development time by analyzing how scientists characterize a habitat.** Determine what tasks scientists conduct during a search-for-life rover mission, what information is used in these tasks, the time devoted to each analysis activity, and the reliability of the conclusions that may be drawn from these tasks. The expected significance of this objective is that future missions may be developed more quickly based on clear definitions of what analysis tasks will be conducted, what data each analysis requires, how long each analysis takes, and the limitations of the conclusions that may be drawn from each analysis.

2. **Reduce mission development risk by identifying mission-critical and problematic analysis tasks.** Evaluate and compare the observations made by control-room scientists during a search-for-life rover mission with observations of the same area made by scientists in the field and laboratory. Determine which control room conclusions were accurate with appropriate confidence levels and which conclusions were false or made with inappropriate confidence levels. The expected significance of this objective is the identification of analysis tasks based on rover-collected data that could mislead scientists' conclusions regarding evidence of life on Mars.
3. **Increase science return from the data by analyzing long-traverse science collection strategies.** Assist in determining the effect of alternative rover data collection strategies during a long (1 km +) autonomous rover traverse, comparing the effect of collecting data samples at regular intervals, collecting data samples at locations specified by scientists in advance, and collecting data samples at locations determined by the rover's autonomous subsystems. The expected significance of this objective is the validation that the proposed analysis techniques can help to optimize rover exploration strategies to ensure the greatest scientific return for each rover mission.
4. **Increase data return by developing science information interface strategies that improve analysis effectiveness.** Refine an existing science information interface to specifically support scientists in making accurate and efficient observations and conclusions. The expected significance of this objective is the development of improved science operations software and the quantification of science performance gains resulting from the analysis and development supported in this proposal.

2.6.3 Timeline

The timeline of the original proposed project is outlined below. According to this timeline all of the tasks up to task 2.7 should be complete by the date of this report. Nearly all of these tasks have been completed. The exception is that two analysis tasks from the 2004 LITA expedition are still ongoing because their results must be compared with the results from the 2005 LITA expedition to determine their significance. There are also several rock and soil samples from the 2004 expedition that are currently undergoing commercial laboratory analysis.



2004-2005

1.1 ***Participate in the Atacama field campaign****, two weeks of operations in Pittsburgh, collect audio and video tape recordings, as well as computer logging data indicating which data was used by whom at what time. Send daily reports of science observations and conclusions to the field for confirmation by the instrument scientists, including sample collections. (All)* This task was successfully completed. Each day of operations 2 members of the Iowa team were present in the control room throughout the test. Each scientist wore a lapel microphone and six cameras were mounted in the ceiling to record the position of all the scientists. Each day of the mission we interrupted the scientists to ask them to make 3 to 5 specific, testable hypotheses. These hypotheses were then tested in the field either immediately by team members in the field, or afterwards when members of our team visited the Atacama.

1.2 ***Create transcripts of the audio recordings from science team leaders.*** *Use video recordings and data logs to trace which teams were working together at each point in the mission and to which data they were referring. Analyze the person-hours that accompanied each individual analysis task in order to find analysis bottlenecks among the scientists. (Thomas)* This task was successfully completed. Full transcripts of the science team activities were collected and reported in the literature (e.g., Pudenz, Glasgow, Thomas, et al., "Searching for a Quantitative Proxy for Rover Science Effectiveness.").

1.3 ***Review transcript analysis, task analysis and interpretation network for domain-specific accuracy.*** *Resolve controversial observations from laboratory studies of samples from the field. Trace any analysis errors revealed in the field back through the transcript to find the origin of each interpretation error. (Thomas, Cabrol, Anderson, Grin)* Complete. As mentioned in the summary section, there was also some difficulty in getting the scientists to unanimously commit to a particular interpretation for different environmental features. Consequently, we found few direct contradictions to scientific findings in the control room. Those that we found, however, have been thoroughly investigated and are reported in Jamie Nakamoto's 2006 Masters Thesis.

1.4 ***Study how the scientists interacted with the information tools and other artifacts in the control room****. Study how the tools meet or did not meet the scientists needs throughout the mission. Determine if training, tools or equipment could have prevented each error and whether the error might be a factor in past or future Mars missions. Create recommendations regarding the information tools provided to the science team and determine what features or functions would be useful in future missions. (Thomas, Coppin).* Complete. The first principle recommendation from the first year of study involved the analysis of the EventScope interface (included in the first annual report as Appendix 1) which determined the substantial improvement made when several refinements were made to the system after the first week of operations. The second recommendation involved eliminating the collection and processing of the stereoscopic images, which did not seem to be used heavily in the science analysis (e.g., Glasgow, Pudenz, Thomas, Cabrol, Coppin and Wettergreen (2005), Observations of a Science Team during an Advanced Planetary Rover Prototype Field Test). This controversial conclusion was contested by some of the scientists, but ultimately the engineers down-graded the priority of the stereo-imaging. The analysis ultimately led to the development of the Directed Autonomy model, which is described in Erin Pudenz's Master's Thesis at the University of Iowa and two journal articles, which are currently in review.

1.5 **Refine the science information systems** (Thomas, Coppin). Complete. Coppin and his team refined the EventScope and website interfaces throughout the mission. Several of these changes were a direct result of the analysis created as part of this project. Most notable among these are the specific benefit observed with adding templates to the programming interface and the creation of a web interface that allowed the scientists to easily access the highest resolution version of the available data.

2005-2006

2.1 ***Participate in the 2005 field campaign****, two to three weeks of science operations at NASA Ames. Video and audio record scientist observations, as before, following interrupt interview protocol, and track information usage with computer-tracking software. (All)* Complete. We participated in the LITA mission operations room. Two members of the Iowa team were again present throughout the mission. Based on the observations of the first year, we decided not to transcribe the entire mission, which was very expensive. Instead we are using digital recordings and hope to combine our detailed notes and the fast access to various points of the conversation to substitute for the complete transcripts. This year we also abandoned the interrupt protocol because of the difficulty some scientists faced when asked to commit to a particular testable hypothesis. Instead we have based our analysis on the science reports produced by the science team each day. These reports have proven to be much richer sources of specific observations and hypotheses than the verbal protocols and also have greater validity than comments pulled for casual conversations. We also changed the protocol to include the consistent observation of scientist activity every five minutes throughout the mission operations.

2.2 ***Travel with science team to Atacama.*** *Visit science sites explored by rover and audiotape and record science team's observations of the sites for comparison with control room observations. (Thomas, Cabrol, Grin and Anderson)* This year we made two trips to the Atacama Desert in Chile. The first trip was made before the Carnegie

Mellon Team had left the field. Ingrid Ukstins Peate, a University of Iowa geologist, traveled with Thomas to 4 of the rover test sites. Together Thomas and Ukstins Peate developed a protocol to test the various comments made in the transcript and in the science summaries, took reference images of nearly all the rover stopping locations, and collected rock and soil samples for laboratory analysis. Thomas and Ukstins Peate returned in early January to study the 3 remaining rover sites. This trip was immediately followed by the full science team and two more Iowa student team members arriving in Chile. The whole team traveled to all the rover sites. We provided books to the science team members that documented all their science reports, and the key images collected from each site. We also highlighted those sites that were of greatest scientific interest because of interpretations made and verified or contradicted by our earlier site visits. For each week of operations we took the science team to two rover locations and asked them to reflect on approximately 7 features of the locales drawn from the science summaries. We believe that these records and the video and audio recordings made in the field are the first records of a scientist reflecting on their own observations made through a robot.

2.3 ***Repeat task-analysis and observation-map analysis to determine if significant changes in the scientist performance could be attributed to improvements in the tools the scientists used.*** *(Thomas)* Complete. The new Directed Autonomy model provides an opportunity to quantitatively describe the behavior transitions made by the scientists as they completed their analysis, which clearly demonstrated changes before and after new data arrived during the daily science cycle. The large number of confounding variables caused by the large-scale field test restricts more detailed analysis of the affect of specific information system changes on individual behaviors. To counter this difficulty, we have chosen to develop more focused tests in the third year, rather than pursue another large-scale field experiment.

2.4 ***Validate transcript analysis to verify domain-specific information.*** *Resolve controversial findings with results of laboratory analysis. Trace any errors revealed in the field exercise back to the original data that provided them. (Thomas, Cabrol, Grin and Anderson)* Complete. Ultimately we pursued three different approaches. The first was to analyze the daily science summaries, collect a series of over 1000 specific observations, tie these to specific locales, and drive to each locale with a geologist (Ingrid Ukstins Peate) and determine whether each of the statements was true or false. The second approach was to select approximately 10 scientifically interesting observations at six different locations in the Atacama Desert and drive the full geology team to each of these locations and ask them to interpret for themselves whether or not the statements were correct. The third approach was to select three "critical incidents" and conduct an investigation in the manner in which a traffic accident or crash investigation is conducted. Each approach had different strengths and weaknesses and provided a different perspective on the mission.

2.5 ***Create recommendations regarding the information tools provided to the science team*** *and determine what features or functions would be useful in future missions. (Thomas, Coppin)* Complete. We developed a mathematical model to describe the relationship between the scientific targets the scientists wish to observe and the necessary resolution for the camera used to make the observations. This work forms the basis for Justin Glasgow's Masters thesis at the University of Iowa and is currently in peer review.

2.6 ***Prepare software for the third-year field test****. (Thomas, Coppin)*. Ongoing. The preliminary design document has been developed based on the semiotic framework proposed. Coppin is currently developing prototypes of the new interface.

2.7 ***Identify and prepare logistics for third-year field test.*** *(Cabrol, Grin and Anderson)*. This task has been modified to include developing the protocols and datasets necessary to begin preliminary tests of the semiotic approach to field test evaluation. The collection of these datasets may require several field trips to interesting geologic sites. Where possible, the collection of these datasets will be tied to other expeditions to minize the cost of the collection while maximizing the breadth of the dataset.

2006-2007

3.1 ***Travel to third-year test site.*** *Collect a data sequence simulating a rover moving about the remote terrain. (Thomas, Anderson, Cabrol, Grin)*. We propose modifying this task to include the collection of the data sets for the semiotic network test.

3.2 ***Install field data on three version of science interface software.*** *Recruit three science teams to participate in three sessions, one with each software package, separated by a period of at least one month. Record audio, video and data interaction during the experience*. We propose modifying this task to include the iterative testing, refinement and exploration of the new semiotic network interface.

3.3 ***Analyze transcripts to determine what data helped the geologists make different observations.*** *Compare these data sources to data sources that are more traditionally build into a rover. (All).* We propose modifying this task to include the analysis of the success of the new interface approach.

Challenges and Opportunities:

Last year we discovered that objective truth is more nebulous than we had expected. The fundamental logic of this proposal was that we would compare the observations of the scientists with the "true" observations in the field and laboratory. The difference between the rover-mediated observations and the field observations would represent the "error" which we would then design to correct. Several factors seemed to work against this approach:

1. The scientists often expressed their ideas as a range of possibilities. Sometimes this range was so large that it precluded very little. In some cases it would be nearly impossible to make a field observation that contradicted the science summary.
2. Sometimes the meaning of a word used by one scientist differed from the meaning used by another scientist. Consequently if a scientist in the control room reported that there was a desert pavement habitat, for example, his or her notion of what constituted a desert pavement habitat might be different from a scientist in the field. Consequently, although the difference of opinion might be recorded as an "error," it was merely a difference of meaning.
3. In the field several scientists could look at the same feature and come to different conclusions. For example, at one site the scientists were asked whether there was a paucity of white rocks. Three scientists said that there was a paucity of white rocks. Another, who happened to be very interested in white rocks, and had been moving from one to the next, examining each in turn, reported that there was no paucity of white rocks. Apparently "truth" depends on one's frame of reference.
4. In some cases, when the observation in the field seemed to contradict an observation made in the control room, the scientists searched for a reason for the difference and, instead of reporting that there was a difference in the observations, sought to explain why the differences were there. It is very difficult to separate rationalization from objective observation in these cases.

We have begun to develop a strategy that we think will address this challenge. Rather than emphasize the outcome, or truth, of a statement, we will focus on the details of the process that leads to the outcome. In order to study the process, we need to propose a model of the process. We have settled on a concept called a semiotic network, which proposes that scientists build their understanding of an environment by constructing a network of ideas. We developed a network framework by sitting with two geologists and developing a tree of observations that would lead a geologist to conclusions about the overall geologic history of an environment. We discovered that at least two parallel trees are required: one for nomenclature (going from general to specific rock types) and one for observations (branching towards different categories of observation, such as color, composition, grain shape and size, rock shape, and composing elements). For a specific rock or outcrop, we theorize that a "good" geologist would make a number of observations and realizations about the rock or outcrop, gradually proceeding through an investigation until the history of the rock is resolved as well as it might be with the instruments available. One might be able to resolve the different contributing observations and logical conclusions unambiguously in a laboratory. As the network evolves, different geologists might interpret the "signs" differently. Some of these interpretations, although reasonable, may be different. Other interpretations or observations may simply be invalid. By structuring the flow of logic that a geologist might pursue about a rock (but permitting many paths through this logical flow), we provide the opportunity to quantify the logic of the process while minimally intruding on the order in which the geologist considers the different features of the rock. Unfortunately, the technique still forces the geologist into a particular paradigm. It remains to be seen whether this is an uncomfortable fit and whether or not geologists feel that the measure of their performance matches their intuition of their performance.

To address these questions, we are devising a series of experiments in which geologists will be presented with sample data (starting with images only) of different rocks. They will be asked to develop a network describing their reasoning about the information in the image using a software interface developed by Coppin. We will score their results and ask them whether the score and the network. Our experimental hypothesis is that both the score and the network will accurately reflect the geologists' thinking about the image. If the experiment is successful, then we will have developed a sensitive instrument for measuring geological performance that may be generalized to other research fields. We can then use this understanding of the scientific thinking process to develop other tools and technologies to support scientists in their work.

Publications and Presentations resulting from this effort

1. Glasgow, J., G. Thomas, E. Pudenz, N. Cabrol, D. Wettergreen, P. Coppin (submitted to IEEE SMC-A, April 2006), “Optimizing Information Value: Improving Rover Sensor Data Collection.”
2. Pudenz, E., G. Thomas and J. Glasgow (submitted July 2006) Directed Autonomy: A new model for a new mobile robot technology, Journal of Human Factors.
3. Pudenz, E., G. Thomas and J. Glasgow (submitted August 2006), “Refining Directed Autonomy: An Empirical Analysis of a Predictive Mode”, Journal of Human Factors.
4. Glasgow, J., G. Thomas, E. Pudenz, N. Cabrol, D. Wettergreen, P. Coppin, “Panoramic Image Information Utility for Mobile Robot Exploration,” IEEE SMC Conference 2006
5. Pudenz, E., Glasgow, J., Thomas, G., Coppin, P., Wettergreen, D., Cabrol, N. Searching for a Quantitative Proxy for Rover Science Effectiveness, Proceedings of the 2006 Conference on Human-Robot Interaction, March 2-4, 2006, Salt Lake City, Utah. (30% acceptance rate)
6. Pudenz, E., Glasgow, J., Thomas, G., Coppin, P., Wettergreen, D., Cabrol, N. Searching for a Quantitative Proxy for Rover Science Effectiveness, Proceedings of the 2006 Conference on Human-Robot Interaction, March 2-4, 2006, Salt Lake City, Utah.
7. Glasgow, Justin, Erin Pudenz, Geb Thomas, Nathalie Cabrol, Peter Coppin and David Wettergreen (2005), Observations of a Science Team during an Advanced Planetary Rover Prototype Field Test, Ro-MAN Conference, August 13-15, Nashville, TN, 2005.
8. Thomas, G., Coppin, P., Cabrol, N., Wettergreen, D., Pudenz, E., Glasgow, J. (2005), Collaborative Virtual Environments for Control of Planetary Exploration Rovers, Special Session on Human Robot Interaction, Human Computer Interaction / Virtual Reality Conference 2005, July 22-27, 2005, Las Vegas, NV.
9. G. Thomas, Engineering Robotic Geology for Mars Exploration, IIE Annual Conference, March 2004, Houston, TX.
10. Keynote speech, “Science with Robots on Earth, the Moon and Mars,” ASME Student Leadership Training Seminar, Iowa City, IA 9/24/05.

Masters Degrees Completed

Student	Title of MS Thesis	Year
Jamie Nakamoto	Discrepancy investigation protocol: Applications to field and vicarious science	2006
Justin Glasgow	Optimizing information value: Reflections on rover sensor data	2006
Erin Pudenz	Directed Autonomy: A New Model for a New Mobile Robot Technology	2006